CERTIFICATE OF TRANSMISSION

I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being submitted *via* the USPTO EFS Filing System on the date shown below to **Mail Stop Appeal Brief - Patents**, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Date: May 5, 2008 /Jessica Sexton/
Jessica Sexton

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of:			
Applicant(s):	Eric Brill, *et al.*	Examiner:	Jude Jean Gilles
Serial No:	10/670,681	Art Unit:	2143
Filing Date:	September 25, 2003		

Title: SYSTEMS AND METHODS FOR CLIENT-BASED WEB CRAWLING

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

APPEAL BRIEF

Dear Sir:

Applicant submits this brief in connection with an appeal of the above-identified patent application. Payment is being submitted via credit card in connection with all fees due regarding this appeal brief. In the event any additional fees may be due and/or are not covered by the credit card, the Commissioner is authorized to charge such fees to Deposit Account No. 50-1063 [MSFTP475US].

I. Real Party in Interest (37 C.F.R. §41.37(c)(1)(i))

The real party in interest in the present appeal is Microsoft Corporation, the assignee of the present application.

II. Related Appeals and Interferences (37 C.F.R. §41.37(c)(1)(ii))

Appellants, appellants' legal representative, and/or the assignee of the present application are not aware of any appeals or interferences which may be related to, will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. Status of Claims (37 C.F.R. §41.37(c)(1)(iii))

Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112, 114-116 stand rejected by the Examiner. The rejection of claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112, 114-116 is being appealed.

IV. Status of Amendments (37 C.F.R. §41.37(c)(1)(iv))

No amendments have been submitted after the Final Office Action.

V. Summary of Claimed Subject Matter (37 C.F.R. §41.37(c)(1)(v))

A. Independent claim 1

Independent claim 1 recites a data analysis system, comprising: a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained *via* a communication system and a second component that coordinates a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system; the second data set is utilized to refine the first data set, wherein refining the first data set comprises adding unknown information to the first data set when new information is received from the distributed source via the second data set or updating existing information in the first data set when changes have occurred in the contents of the web page information as indicated by the second data set (*See. e.g* Figs 1, 2, 8 and corresponding text at pg. 7 line 17 to pg. 8 line 24 and pg. 20, line 8 to pg. 21 line 6).

B. Independent claim 37

Independent claim 37 recites a method for facilitating data analysis, comprising: generating a first data set relating to a second data set obtained from web pages interactive with a server of a communication system; receiving a third data set from at least one distributed resource comprising a client of the server that is interactive with the communication system; the third data set comprising web page related information generated by the distributed resource; and refining the second data set to reflect information obtained from the third data set, by: adding unknown information to the second data set when new information is received from the distributed source *via* the third data set; updating existing information in the second data set when changes have occurred as indicated by the third data set; and passing status information to the distributed resource through one or more indicators after information from the third data set has been analyzed(*See. e.g* Figs 3, 8 and corresponding text at pg. 8 line 25 to pg. 11 line 31 and pg. 20, line 8 to pg. 21 line 6).

C. Independent claim 57

Independent claim 57 recites a data analysis system, comprising: means for generating at least one first data set from a server of communication system; means for receiving and coordinating at least one second data set from at least one client which interacts with the server of the communication system (Fig 5, component 510); and means for refining the first data set utilizing at least one second data set, wherein refining the first data set comprises the at least one of adding unknown information to the first data set when new information is received from the client via the second data set and updating existing information in the first data set when changes have occurred in the web page as indicated by the second data set (*See. e.g* Figs 3, 5 and corresponding text at pg. 8 line 25 to pg. 11 line 31 and pg. 14 line 25 to pg. 16, line 18).

The aforementioned means for limitations are identified as claim elements subject to the provisions of 35 U.S.C. §112 ¶6. The corresponding structures are identified with reference to the specification and drawings in the parentheticals above corresponding to those claim limitations.

D. Independent claim 61

Independent claim 61 recites a data analysis system, comprising: a first component associated with at least one client of a distributed web crawling system that generates web page information from at least one visited web site for utilization in the distributed web crawling system; and a second component associated with a server that receives the web page information transmitted by the first component *via* a communication system, wherein the first component receives a set of data from the second component to utilize in the generation of the web page information comprising at least comparison data based on the visited web page and the received set of data (*See. e.g* Fig 5 and corresponding text at pg. 14 line 25 to pg. 16, line 18).

E. Independent claim 92

Independent claim 92 recites a method for facilitating data analysis, comprising: compiling a first data set derived from accessing web pages *via* a client of a communication system; transmitting, selectively, the first data set to an entity comprising at least a server of a distributed crawling system that is interactive with the communication system; receiving a representation of a second data set compiled by the server of the web crawler; the second data set relating to at least one web page from the communication system; and utilizing the second data set to control which web pages to visit to compile the first data set (See. *e.g* Fig 8, and corresponding text at pg. 20, line 8 to pg. 21 line 6).

F. Independent claim 114

Independent claim 114 recites A computer readable medium having stored thereon computer executable components comprising: a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained *via* a communication system; and a second component that coordinates a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system; the second data set is utilized to refine the first data set, wherein refining the first data set comprises adding unknown information to the first data set when new information is received from the distributed source via the second data set and updating existing information in the

first data set when changes have occurred in the contents of the web page information as indicated by the second data set (See. *e.g* Fig 3, 8 and corresponding text at pg. 8 line 25 to pg. 11 line 31 and pg. 20, line 8 to pg. 21 line 6).

.

VI. Grounds of Rejection to be Reviewed (37 C.F.R. §41.37(c)(1)(vi))

A. Whether claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116 are unpatentable under 35 U.S.C. §103(a) over Bailey *et al*, (U.S. 20060167864) in view of Albion *et al*. (U.S. 20040240388).

VII. Argument (37 C.F.R. §41.37(c)(1)(vii))

A. Rejection of Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116 Under 35 U.S.C. §103(a)

Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116 stand rejected as unpatentable under 35 U.S.C. §103(a) over Bailey *et al*, (U.S. 20060167864) in view of Albion *et al*. (U.S. 20040240388). Reversal of this rejection is requested for at least the following reasons. Bailey *et al*. and Albion *et al*. alone or in combination fail to teach or suggest all features set forth in the subject claims.

Appellants' claimed subject matter relates to data analysis, and systems and methods for obtaining information from a networked system utilizing a distributed web crawler. Information gathered by a server's web crawler is compared to data retrieved by clients of the server to update the crawler's data. In particular, independent claim 1 recites *a data analysis system, comprising: a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained via a communication system; and a second component that coordinates* ***a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system***. Independent claims 37 further recites *refining the second data set to reflect information obtained from the third data set* ***by adding unknown information to the second data set when new information is received from the distributed***

source via the third data set. Independent claims 57, 61, 92 and 113 recite similar features. Bailey *et al.* and Albion *et al.* are silent regarding such novel features.

Bailey *et al.* relates to a search engine system for locating web pages with product offerings. At page 3 of the Final Office Action, the Examiner contends that Bailey *et al.* discloses such novel features of applicants' claimed invention. Appellants' representative avers to the contrary. In accordance with the subject invention, a server hosts a web crawler that searches a communication network such as the Internet for other servers hosting web pages, gathers information about these web pages and compiles them for utilizing with a web page search engine (*See* applicants' Fig.1 and Fig.2). The server then sends a representation of this web page information to a client of the server. When the client accesses that particular web page or detects web pages that are unknown to the server, the client compiles changes/status and/or new information about the known and unknown web pages. This information is then transmitted to the server, which utilizes the information to update its original crawler web page data to reflect a new web page or change of contents in a known web page. At the cited portions, Bailey *et al.* discloses a web server application that processes user requests to query and make purchases from a catalog, *via* the internet 120. The web server records the user transactions within a query log. Further, Bailey *et al.* discloses the Product Spider database that has product scores and category ranking information about independent web sites unaffiliated with the host web site, that offer products for sale. When updating the database, URL's of the existing database are submitted to the second crawling stage, updated, duplicate submissions are detected and removed. However, the cited document is silent regarding utilizing web page information communicated by a client of the distributed web crawler system to update its original crawler web page data to reflect a new web page or change of contents in a known web page. For example, Bailey *et al.* teaches a conventional web crawler implemented by a server but does not teach or suggest that the web crawler 160 is updated with inputs from the clients 110 (*See* Bailey *et al.* Fig. 1 and paragraph [0037]) Thus Bailey *et al.* does not disclose a distributed web crawler wherein a client updates web pages associated with a server of the distributed system as recited by the subject claims.

Albion *et al.* relates to dynamic assignment of timers in a network transport engine that provides a connection between two applications running on different system interconnected via the network. Timer logic includes a counter, a crawler, a memory and a list of available timers.

The crawler processes client request and keeps note of the timer information in the timer list located in the memory. Upon a request from the client, timers are allocated, de-allocated and restarted. Accordingly the timer list is updated. However, the second data set from the client is not web page information communicated by a client of the distributed web crawler system as recited by the subject claims. Thus, crawler 204 of Fig. 2 in Albion is a component that manages timers accessed by clients rather than a component that provides unknown information or updates information when changes have occurred in the contents of the web page information as recited in the subject independent claims (*See e.g.,* Albion paragraph [0018]). Therefore, it is concluded that Albion *et al.* is silent regarding *refining the second data set to reflect information obtained from the third data set* ***by adding unknown information to the second data set when new information is received from the distributed source via the third data set*** as recited by the subject claims.

By distributing the web crawler functionality among the search server and its clients, the server utilizes the clients to obtain information from web page servers to facilitate in refining its own information. This helps in providing a more up-to-date, robust and spoof-proof data set from which a search engine can utilize data.

In view of at least the foregoing, it is readily apparent that both Bailey *et al.* and Albion *et al.* fail to teach or suggest all limitations of the claimed invention. Accordingly, it is respectfully requested that rejection of independent claims 1, 37, 57, 61, 92 and 113 (and the claims that depend there from) be reversed.

B. Conclusion

For at least the above reasons, the claims currently under consideration are believed to be patentable over the cited references. Accordingly, it is respectfully requested that the rejections of claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112, 114-116 be reversed.

If any additional fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP475US].

Respectfully submitted,
AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/
Himanshu S. Amin
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP
24th Floor, National City Center
1900 East 9th Street
Cleveland, Ohio 44114
Telephone: (216) 696-8730
Facsimile: (216) 696-8731

VIII. Claims Appendix (37 C.F.R. §41.37(c)(1)(viii))

1. A data analysis system, comprising:

a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained *via* a communication system; and

a second component that coordinates a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system; the second data set is utilized to refine the first data set, wherein refining the first data set comprises at least one of adding unknown information to the first data set when new information is received from the distributed source via the second data set and updating existing information in the first data set when changes have occurred in the contents of the web page information as indicated by the second data set.

2. The system of claim 1, the first component comprising an internet web crawler.

3. The system of claim 1, the first component comprising an intranet web crawler.

4. The system of claim 1, the second component further utilized to optimize reception of data from the distributed resources.

5. The system of claim 1, the second component provides a scheduling function to control reception of the second data set from the at least one distributed resource.

6. The system of claim 1, the second component utilized to facilitate communication traffic reduction *via* the communication system by employing a proper set of weak indicator functions representative of the first data set.

7. The system of claim 6, the second component further utilized to randomly select and transmit a weak indicator function selected from the proper set of weak indicator functions to at least one of the distributed resources.

8. (Cancelled)

9. The system of claim 1, the second component further utilized to generate status information about data related to the first data set; the status information transmitted to at least one distributed resource.

10. The system of claim 9, the status information comprising, at least in part, a freshness flag to indicate freshness of information related to the first data set.

11. The system of claim 9, the status information comprising, at least in part, a hash of contents of information related to the first data set.

12. The system of claim 9, the status information comprising, at least in part, a copy of information of the first data set.

13. The system of claim 1, the communication system comprising an internet.

14. The system of claim 1, the communication system comprising a world wide web.

15. The system of claim 1, the communication system comprising an intranet.

16. The system of claim 15, the intranet comprising a local area network.

17. The system of claim 15, the intranet comprising a wide area network.

18. (Cancelled)

19. The system of claim 1, the distributed resources comprising trusted entities interactive with the communication system and the second component.

20. The system of claim 1, the first data set comprising internet web page data.

21. The system of claim 1, the first data set comprising intranet web page data.

22. (Cancelled)

23. The system of claim 1, the second data set comprising, at least in part, a hash of contents of at least one web page.

24. The system of claim 1, the second data set comprising, at least in part, a Uniform Resource Locator (URL) of at least one web page.

25. The system of claim 1, the second data set comprising, at least in part, a time stamp relating to an acquisition time for information about at least one web page.

26. The system of claim 1, the second data set comprising, at least in part, a delta indication of the changes to contents of the at least one web page.

27. The system of claim 26, the delta indication including, at least in part, a hash of previous contents of a web page and a hash of recent contents of the web page.

28. The system of claim 1, the second data set comprising, at least in part, a status indication of changes to contents of at least one web page.

29. The system of claim 28, the status indication including, at least in part, a percentage relating to an amount of change of contents of a web page.

30. The system of claim 28, the status indication including, at least in part, a significance indicator to signify importance of changes in contents of a web page.

31. The system of claim 1, the second data set comprising internet web page data.

32. The system of claim 1, the second data set comprising intranet web page data.

33. The system of claim 1, the second data set comprising data compiled utilizing at least one weak indicator function randomly selected from a set of weak indicator functions; the set of weak indicator functions representative of the first data set.

34. The system of claim 1, further comprising a search component to accept at least one search query and generate at least one search reply having at least a portion of the first data set represented by information embedded in the search reply.

35. The system of claim 1, further comprising a web page server component to construct web pages having at least a portion of the first data set represented by information embedded in at least one link found on at least one constructed web page.

36. The system of claim 1, further comprising a storage component to store the first data set.

37. A method for facilitating data analysis, comprising:

generating a first data set relating to a second data set obtained from web pages interactive with a server of a communication system;

receiving a third data set from at least one distributed resource comprising a client of the server that is interactive with the communication system; the third data set comprising web page related information generated by the distributed resource; and

refining the second data set to reflect information obtained from the third data set, by

adding unknown information to the second data set when new information is received from the distributed source *via* the third data set;

updating existing information in the second data set when changes have occurred as indicated by the third data set; and

passing status information to the distributed resource through one or more indicators after information from the third data set has been analyzed.

38. The method of claim 37, the first data set comprising a representation of the second data set.

39. The method of claim 38, the representation of the second data set comprising, at least in part, a hash of contents of at least one web page contained in the second data set.

40. The method of claim 38, the representation of the second data set comprising, at least in part, a status indication of at least one web page contained in the second data set.

41. The method of claim 40, the status indication comprising a freshness flag to indicate if the web page information is current.

42. The method of claim 37, the first data set comprising a copy of the second data set.

43. The method of claim 37, the second data set comprising web page information compiled by a web crawler.

44. The method of claim 37, the third data set comprising web page information based upon client accessed web page information on the communication system.

45. (Cancelled)

46. The method of claim 37, the communication system comprising an internet.

47. The method of claim 37, the communication system comprising an intranet.

48. (Cancelled)

49. The method of claim 37, further including:

transmitting the first data set to at least one distributed resource that is interactive with the communication system making the first data set available to be utilized by the distributed resource to generate the third data set.

50. The method of claim 38, further including:

generating a set of weak indicator functions to represent the second data set; and

selecting random weak indicator functions from the set of weak indicator functions to transmit to the distributed resources as the first data set.

51. The method of claim 50, the set of weak indicator functions comprising a proper set of weak indicator functions such that a non-zero probability exists that a randomly selected weak indicator function can identify a new web page.

52. The method of claim 50, generating a set of weak indicator functions comprising:

providing a dictionary representative of the second data set;

partitioning randomly the dictionary into non-overlapping subdictionaries; and

creating a function where $I(x) = 1$ if and only if at least one subdictionary's weak indicator function is equal to one.

53. The method of claim 37, further including:

comparing the third data set to the second data set to reveal spoof data included in the second data set.

54. The method of claim 37, further including:

optimizing reception of at least one third data set through scheduling of the distributed resources.

55. The method of claim 37, further including:

receiving a web page search query from at least one distributed resource;

generating a web search results page in response to the web page search query from the distributed resource;

embedding portions of the first data set in links found on the web search results page; and

transmitting the web search results page as a representation of at least a portion of the second data set to the distributed resource.

56. The method of claim 37, further including:

constructing a web page utilizing at least a portion of the first data set to embed information about links found in the web page; and

transmitting the web page to disseminate the first data set to at least one distributed resource.

57. A data analysis system, comprising:

means for generating at least one first data set from a server of communication system;

means for receiving and coordinating at least one second data set from at least one client which interacts with the server of the communication system; and

means for refining the first data set utilizing at least one second data set, wherein refining the first data set comprises the at least one of adding unknown information to the first data set when new information is received from the client via the second data set and updating existing information in the first data set when changes have occurred in the web page as indicated by the second data set.

58. The system of claim 57, the means for generating at least one first data set including a web crawler.

59. The system of claim 58, the first data set comprising data relating to web pages obtained by the web crawler.

60. The system of claim 57, the second data set comprising web page comparison data compiled by the at least one client and based, at least in part, upon representative data of the first data set.

61. A data analysis system, comprising:

a first component associated with at least one client of a distributed web crawling system that generates web page information from at least one visited web site for utilization in the distributed web crawling system; and

a second component associated with a server that receives the web page information transmitted by the first component *via* a communication system, wherein the first component receives a set of data from the second component to utilize in the generation of the web page information comprising at least comparison data based on the visited web page and the received set of data.

62. The system of claim 61, the first component providing at least one time stamp relevant to a time of acquisition of data utilized in the generation of the web page information.

63. The system of claim 61, the first component receiving a set of embedded web crawler data from at least one search result page to utilize in the generation of the web page information.

64. The system of claim 61, the first component receiving a set of embedded web crawler data from at least one web page to utilize in the generation of the web page information.

65. The system of claim 61, the first component further operational to obtain web page data indirectly *via* at least one other client of the distributed crawler system to provide a gateway to the second component to substantially reduce traffic flow to the second component.

66. (Cancelled)

67. The system of claim 61, the generated web page information comprising, at least in part, a status indication of changes to contents of at least one web page.

68. The system of claim 67, the status indication including, at least in part, a percentage relating to an amount of change of contents of a web page.

69. The system of claim 67, the status indication including, at least in part, a significance indicator to signify importance of changes in contents of a web page.

70. The system of claim 61, at least a portion of the generated web page information made available for peer-to-peer client transmission *via* the communication system.

71. The system of claim 61, the generated web page information compiled utilizing a randomly selected weak indicator function from a proper set of weak indicator functions that represent web page data compiled by a web crawler.

72. The system of claim 61, the communication system comprising an internet.

73. The system of claim 61, the communication system comprising an intranet.

74. The system of claim 61, further comprising a storage component to store the web page information.

75. The system of claim 61, further comprising a notification component that determines when and if the generated web page information is to be communicated *via* the communication system.

76. The system of claim 75, the notification component receiving scheduling information from the second component; the scheduling information relating to obtaining and transmitting the generated web page information.

77. (Cancelled)

78. The system of claim 61, the first component utilizing web search servers outside of the distributed web crawling system to retrieve data unknown to the second component.

79. The system of claim 61, the first component making the comparison data discretionarily available to the second component *via* the communication system.

80. The system of claim 61, the comparison data including, at least in part, at least one Uniform Resource Locator (URL) of at least one web page.

81. The system of claim 61, the comparison data including, at least in part, a hash of contents of at least one web page representative of a recent web site visit.

82. The system of claim 61, the comparison data including, at least in part, a delta indication of contents of at least one web page.

83. The system of claim 82, the delta indication including, at least in part, a hash of previous contents of a web page and a hash of recent contents of the web page.

84. The system of claim 61, the second component comprising a server of the distributed crawling system.

85. The system of claim 61, the second component comprising a client of the distributed crawling system.

86. The system of claim 61, the generated web page information comprising data unknown to the second component.

87. The system of claim 61, at least a portion of the received set of data made available for peer-to-peer client transmission *via* the communication system.

88. The system of claim 61, the received set of data comprising a dictionary for data compiled by a web crawler.

89. The system of claim 61 the received set of data comprising a representation of data compiled by a web crawler; the representation of data generated by utilizing a weak indicator function.

90. The system of claim 61, the received set of data comprising a copy of data compiled by a web crawler.

91. The system of claim 61, further comprising a storage component to store the set of data received from the second component.

92. A method for facilitating data analysis, comprising:

compiling a first data set derived from accessing web pages *via* a client of a communication system;

transmitting, selectively, the first data set to an entity comprising at least a server of a distributed crawling system that is interactive with the communication system;

receiving a representation of a second data set compiled by the server of the web crawler; the second data set relating to at least one web page from the communication system; and

utilizing the second data set to control which web pages to visit to compile the first data set.

93. (Cancelled)

94. (Cancelled)

95. The method of claim 92, the first data set comprising, at least in part, a uniform resource locator (URL) for at least one web page.

96. The method of claim 92, the first data set comprising, at least in part, a hash of contents of at least one web page.

97. The method of claim 92, selectively transmitting based upon time of day.

98. The method of claim 92, selectively transmitting based upon priority of at least one web page.

99. The method of claim 92, selectively transmitting based upon percentage of content change of at least one web page.

100. The method of claim 92, selectively transmitting based upon identifying at least one new web page.

101. (Cancelled)

102. The method of claim 92, receiving the representation of the second data set is accomplished *via* reception of a web page with embedded information derived from the second data set and generated by a web page hosting server with access to the second data set.

103. The method of claim 92, receiving the representation of the second data set is accomplished *via* reception of a search results page with embedded information derived from the second data set and generated in response to a query transmitted to a search server having access to the second data set.

104. (Cancelled)

105. The method of claim 92, further comprising:

determining when to transmit the first data set *via* the communication system based upon the second data set.

106. The method of claim 105, the second data set containing a freshness indicator to indicate when its data is stale and requires updating *via* the first data set.

107. The method of claim 105, the second data set containing a schedule for when the first data set is to be transmitted.

108. The method of claim 92, further comprising:

comparing at least a portion of the second data set with at least a portion of information obtained *via* accessing web pages to create comparison data; and

generating a representation of the comparison data to derive the first data set.

109. The method of claim 108, the first data set comprising data unknown to the second data set.

110. The method of claim 109, the unknown data comprising only unknown data derived from at least one search results page from a search server outside of the distributed crawling system.

111. The method of claim 108, the first data set comprising content changes to web pages represented by the second data set.

112. The method of claim 108, the first data set comprising status information relating to web pages represented by the second data set.

113. (Cancelled)

114. A computer readable medium having stored thereon computer executable components comprising:

a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained *via* a communication system; and

a second component that coordinates a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system;

the second data set is utilized to refine the first data set, wherein refining the first data set comprises adding unknown information to the first data set when new information is received from the distributed source via the second data set and updating existing information in the first data set when changes have occurred in the contents of the web page information as indicated by the second data set.

115. A device employing the method of claim 37 comprising at least one selected from the group consisting of a computer, a server, and a handheld electronic device.

116. A device employing the system of claim 1 comprising at least one selected from the group consisting of a computer, a server, and a handheld electronic device.

IX. Evidence Appendix (37 C.F.R. §41.37(c)(1)(ix))

None.

X. Related Proceedings Appendix (37 C.F.R. §41.37(c)(1)(x))

None.